



Objective assessment of surgical skill transfer using non-invasive brain imaging

Arun Nemani¹ · Uwe Kruger¹ · Clairice A. Cooper² · Steven D. Schwaitzberg² · Xavier Intes¹ · Suvranu De¹

Received: 21 March 2018 / Accepted: 12 October 2018 / Published online: 17 October 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

Background Physical and virtual surgical simulators are increasingly being used in training technical surgical skills. However, metrics such as completion time or subjective performance checklists often show poor correlation to transfer of skills into clinical settings. We hypothesize that non-invasive brain imaging can objectively differentiate and classify surgical skill transfer, with higher accuracy than established metrics, for subjects based on motor skill levels.

Study design 18 medical students at University at Buffalo were randomly assigned into control, physical surgical trainer, or virtual trainer groups. Training groups practiced a surgical technical task on respective simulators for 12 consecutive days. To measure skill transfer post-training, all subjects performed the technical task in an ex-vivo environment. Cortical activation was measured using functional near-infrared spectroscopy (fNIRS) in the prefrontal cortex, primary motor cortex, and supplementary motor area, due to their direct impact on motor skill learning.

Results Classification between simulator trained and untrained subjects based on traditional metrics is poor, where misclassification errors range from 20 to 41%. Conversely, fNIRS metrics can successfully classify physical or virtual trained subjects from untrained subjects with misclassification errors of 2.2% and 8.9%, respectively. More importantly, untrained subjects are successfully classified from physical or virtual simulator trained subjects with misclassification errors of 2.7% and 9.1%, respectively.

Conclusion fNIRS metrics are significantly more accurate than current established metrics in classifying different levels of surgical motor skill transfer. Our approach brings robustness, objectivity, and accuracy in validating the effectiveness of future surgical trainers in translating surgical skills to clinically relevant environments.

Keywords Surgical skill assessment · Surgical skill transfer · Brain imaging · Surgical simulators · Surgical training · Functional near-infrared spectroscopy

With mounting concerns about patient safety and the need to have objective measures of surgical technical competence, simulation as a means of surgical training and certification is rapidly gaining ground [1]. The fundamentals of laparoscopic surgery (FLS), which employs a box trainer, and the fundamentals of endoscopic surgery (FES) with a virtual reality-based simulator, have been recently adopted by the

American Board of Surgery as pre-requisites for certification in general surgery [2–10]. However, prior to acceptance, each simulator, real or virtual, must undergo extensive validation and show evidence of successful transfer of technical skills from the simulation environment to the clinical environment [1, 11, 12].

The current standard in assessing successful transfer of skills from the simulation environment to a clinical setting is direct observations by an expert clinician [13] using a checklist such as the objective structured assessment of technical skills, or global operative assessment of laparoscopic surgery [13–15]. Alternative metrics such as task completion time have also been reported for assessing technical skill transfer [16]. Despite the current widespread usage of these generalized rating or completion time-based assessments, there are significant drawbacks to these methods that

Presented at ACS Clinical Congress 2017, San Diego, CA.

✉ Suvranu De
des@rpi.edu

¹ Rensselaer Polytechnic Institute, 110, 8th Street, Troy, NY 12180, USA

² University at Buffalo School of Medicine and Biomedical Sciences, Buffalo, NY 14228, USA

include personnel resource costs, poor interrater reliability between proctors, and poor correlation of learned technical skills from the simulator to outcomes in the operating room [16–18]. These limitations necessitate a need for more objective and analytical methods to assess surgical skill transfer [19, 20].

A promising technique that is objective in determining surgical motor skills is non-invasive brain imaging. Among all the non-invasive brain imaging methods currently available, functional near-infrared spectroscopy (fNIRS) offers the unique features to be portable, non-invasive, non-obtrusive to perform the surgical task, fast and relatively inexpensive [21, 22]. Investigators have used fNIRS to study brain activation responses between surgical experts and novices during the performance of surgical training tasks by measuring the fluctuations of hemodynamics signals, namely changes in concentration of oxygenated and deoxygenated hemoglobin [23–29]. However, these studies are limited in scope as they are subject to signal contamination from superficial tissue, and show no evidence of surgical skill transfer to more clinically relevant environments.

The purpose of this study is to determine if fNIRS can accurately assess motor skill transfer from simulation to *ex vivo* environments for trained and untrained subjects as they perform an established surgical training task. We hypothesize that fNIRS-based metrics can classify different levels of surgical motor skill transfer with more accuracy than established methods. To test this hypothesis, subjects trained on a physical or virtual surgical simulator where they practiced a surgical training task and subsequently performed a surgical transfer task post-training. Based on brain imaging metrics, we then utilize multivariate statistical approaches to objectively differentiate and classify subjects that exhibit successful motor skill transfer.

Methods

Experimental setup

Two different laparoscopic skills trainers were utilized in the study. We utilize the official FLS box trainer as the physical simulator since it is widely used for training laparoscopic skills and is validated for board certification [8, 30, 31]. We use the validated Virtual Basic Laparoscopic Skills Trainer (VBLaST) system, which replicates the FLS pattern cutting task on a computer model with high fidelity [32–37], as the virtual simulator. To perform real-time brain imaging, a fNIRS system (CW6 system, TechEn Inc., MA, USA) was used to deliver infrared light.

In order to measure cortical activation changes during the transfer task, we measure functional activation specifically in the prefrontal cortex (PFC), primary motor cortex

(M1), and the supplementary motor area (SMA), as these cortical regions are directly involved in fine motor skill learning, planning, and execution [28, 38–42]. We design a probe geometry that includes eight infrared illumination sources coupled to 16 long separation detectors and eight short separation detectors. Monte Carlo simulations indicate that our probe design is highly sensitive to functional activation changes in the PFC, M1, and SMA [43]. The distance between the long separation detectors and each corresponding source is within 30–40 mm to ensure specificity to white and gray matter. Furthermore, short separation detectors were placed 8 mm away from each corresponding source to ensure that only superficial tissue layers, such as skin, bone, dura, and pial surfaces are measured. These superficial tissue signals are later regressed during post-processing. A schematic of the probe locations onto the scalp along with probe geometry specifications is shown in Fig. 1.

Subject recruitment and study design

In this IRB approved study, 18 medical student subjects were recruited at University at Buffalo. These subjects had no prior surgical experience and were randomly placed in one of three groups: untrained control ($n=5$), training FLS ($n=7$), and training VBLaST ($n=6$) groups. Only the FLS and VBLaST training groups underwent rigorous training on their respective simulators for 12 consecutive days, completing an average of over 100 pattern cutting trials per subject. The control group did not undergo any training on either simulator. Once training is complete for the FLS and VBLaST groups, all subjects performed a post-test after a 2-week break period to measure surgical skill retention. However, the control group did not undergo training, and performed the post-test and transfer tasks following a 2-week break period immediately after their baseline tests. The post-test consisted of three pattern cutting trials each for all subjects on the FLS and VBLaST simulators. Nemani et al. further details pertinent information on the study design, power calculations for sample sizes, and other experimental design considerations [37].

The transfer task, however, consisted of an FLS pattern cutting task performed on cadaveric abdominal tissue instead of gauze. One cadaveric tissue sample, that consisted of a peritoneum layer with underlying fascia and muscle tissue, was prepared for each subject. While each sample is on average half an inch thick, the peritoneum layer is only a few millimeters in depth. Each sample was circle marked with the same dimensions as the marked circles in the FLS pattern cutting task. Tissue samples were securely placed in the official FLS trainer box. Each subject was then instructed to cut the marked circle on the peritoneal tissue sample and resecting the cut peritoneum section as quickly and accurately as possible without damage to the underlying muscle.

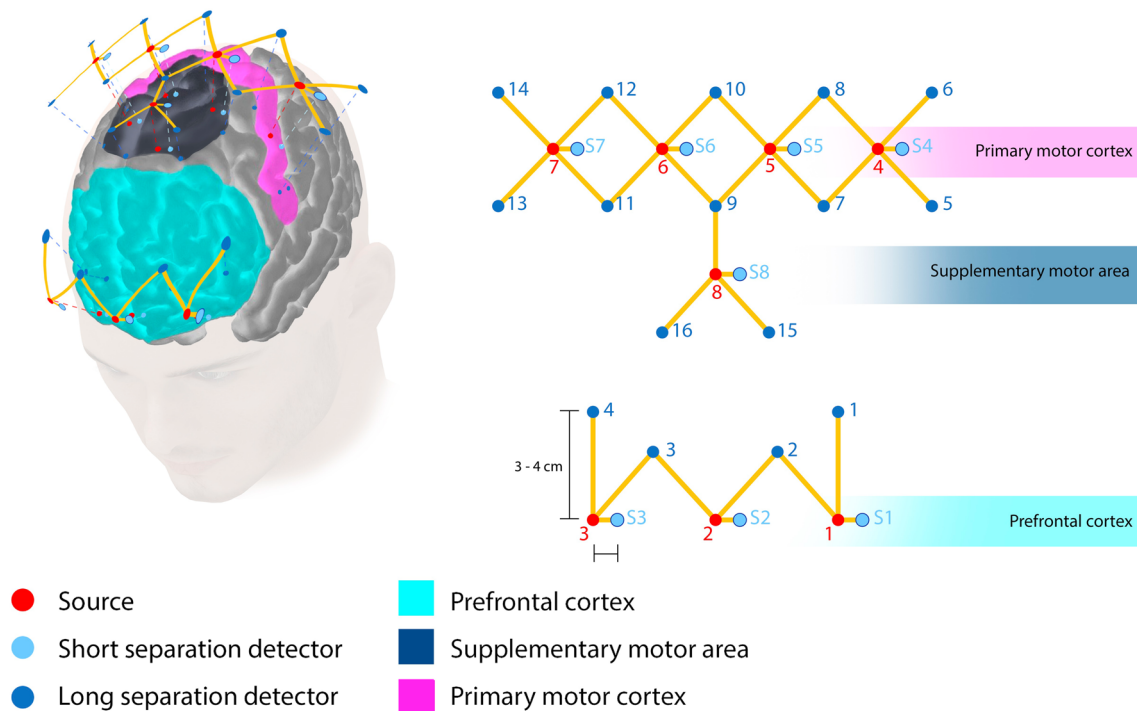


Fig. 1 Infrared probe geometry positioning. Schematic of probe placement projected on cortical locations specific to the PFC, M1, and SMA. Optodes are placed for maximum coverage over the PFC, M1, and SMA. Red dots indicate infrared sources, blue dots indicate long separation detectors, and textured blue dots indicate short separation detectors.

The PFC has three sources (1–3), the M1 has four sources (4–7), and the SMA has one source (8). Each of the sources is connected to their corresponding long and short separation detectors. (Color figure online)

Accredited task performance metrics

Task performance metrics based on time and error have already been established for the FLS and VBLaST simulators. The FLS scoring metrics which are used in Board certification are proprietary, yet were obtained under a non-disclosure agreement with the Society of American Gastrointestinal and Endoscopic Surgeons. The VBLaST pattern cutting score also reproduces the FLS scoring methodology specifically in the virtual environment [36]. As a measure of effectiveness in training, the FLS and VBLaST pattern cutting scores were reported during the post-test to demonstrate that trained subjects significantly outperform untrained subjects. The performance metric for the transfer task was completion time. University policies prohibited video recording of cadaveric tissue and thus no further performance measures could be obtained. Task completion time, with an accuracy of ± 1 s, consisted of the total time (minutes) required to cut and resect the marked peritoneal tissue from the overall tissue sample.

Neuroimaging-based performance metrics

Functional brain imaging using fNIRS was utilized to derive a metric for measuring bimanual surgical skill performance in this study. Prior to data analysis, only measurement channels within the signal qualities between 80 and 140 dB were included. The wavelengths measured at 690 nm and 830 nm, with their corresponding partial path-length factors of 6.4 and 5.8, respectively, were converted to optical density using the modified Beer–Lambert law [44, 45]. Motion artifacts and systemic physiology interference were corrected using low-pass filters and recursive principal component analysis [46–48]. The filtered optical density data were used to derive the change in concentrations of oxy- and deoxy-hemoglobin. To remove signals from superficial tissue layers and increase specificity to cortical tissue hemodynamics, signals from short separation detectors were regressed from long separation detectors [49]. Finally, the corresponding source and detector pairs for each source were averaged over the transfer task completion time. The result is a scalar value for the change in oxy-hemoglobin according to different

brain regions for each participant. All of the fNIRS data processing were completed using the open-source software HOMER2 [46].

Statistical tests and classification approaches

To determine statistical significances between data sets, two tailed Mann–Whitney U tests were utilized within a 95% confidence interval. This statistical test was used for all univariate tests where the type I error is defined as 0.05 for all hypothesis testing cases.

Linear discriminant analysis (LDA) was used for classifying untrained control subjects with either FLS trained or VBLaST trained subjects based on traditional and fNIRS metrics. LDA is an established multivariate classification approach that determines the maximal separation between two different classes based on multivariate metrics [50, 51]. Type I error is defined as 0.05 for all classification models. The quality of classification is reported by misclassification errors (MCE), specifically MCE_{12} and MCE_{21} . MCE_{12} is defined as the probability that a trained subject is misclassified as an untrained subject during the transfer task. Conversely, MCE_{21} is defined as the probability that an untrained subject is misclassified as a trained subject. Theoretically, MCEs of 100% indicate that untrained and trained subjects are identical and indistinguishable, whereas MCEs of 0% indicate that untrained and trained subjects can be classified and differentiated with absolute certainty.

Leave-one-out cross-validation was used to assess how well each classification model can generalize to independent data sets by systematically removing one data point from the data set. Ultimately, cross-validation allows an objective assessment of the robustness of classification models when incorporating potentially new untrained or trained subject data sets. All classification and statistical analysis were completed using Matlab (Mathworks, Natick, MA).

Results

Differentiation and classification of motor skill transfer based on traditional task performance

To investigate whether trained subjects significantly outperform untrained subjects in the ex-vivo environment, first we report transfer task completion times for trained FLS, trained VBLaST, and untrained control subjects. As shown in Fig. 2a, results indicate that both the trained FLS (7.9 ± 3.3 min) and trained VBLAST (12.2 ± 1.8 min) groups completed the transfer task significantly faster than the untrained control group (18.3 ± 3.1 min, $p < 0.05$). While results show that transfer task time can statistically

differentiate trained and untrained subjects during a transfer task, they do not address the accuracy of differentiation.

In this context, LDA-based classification was used to classify trained and untrained subjects based on completion time. Figure 2A shows that classification based on transfer task completion time of trained FLS and untrained control subjects is poor, as shown by high MCEs ($MCE_1 = 20\%$, $MCE_2 = 14\%$). These results indicate that a trained FLS student has a 20% probability of being misclassified as a control subject and an untrained control subject has a 14% probability of misclassified as FLS trained subjects. Cross-validation results, as seen in Fig. 2C, show that 10/12 or 83% of the samples have MCEs less than 5%, indicating that the classification model is valid for potentially future data sets. The same classification approach was applied for the virtual simulator trained (VBLaST) subjects vs untrained control subjects as shown in Fig. 2D. Once again, subject classification based on transfer task completion time is poor, indicated by high MCEs ($MCE_1 = 20\%$, $MCE_2 = 41\%$). Furthermore, cross-validation results show that 8/11 or 72% of the samples have MCEs less than 5%, as shown in Fig. 2e.

Neuroimaging-based metrics for differentiation and classification of motor skill transfer

Due to high MCE encountered in assessing transfer task performance based on task time, we propose subject classification based on fNIRS metrics. Prior to classification, we determine if fNIRS is sensitive to subject cortical activation changes during the transfer task, specifically in the PFC, left medial M1 (LMM1), and the SMA. Results indicate that all simulator trained subjects show no significant differences in all PFC cortical regions compared to control subjects ($p > 0.05$). However, both FLS and VBLaST simulator trained subjects have significantly higher functional activation in the left medial M1 (0.64 ± 0.54 and 0.44 ± 0.18 ΔHbO_2 conc. $\mu\text{M} \cdot \text{mm}$, respectively) compared to untrained control subjects (-0.44 ± 0.72 ΔHbO_2 conc. $\mu\text{M} \cdot \text{mm}$, $p = 0.018$ and $p = 0.004$, respectively). Furthermore, both FLS and VBLaST trained subjects also showed significant increases in functional activation in the SMA (0.42 ± 0.56 and 0.74 ± 0.47 ΔHbO_2 conc. $\mu\text{M} \cdot \text{mm}$, respectively) when compared to untrained control subjects (-0.08 ± 0.22 ΔHbO_2 conc. $\mu\text{M} \cdot \text{mm}$, $p = 0.048$ and $p = 0.009$, respectively). Figure 3A summarizes these descriptive statistics and Fig. 3B shows a visual depiction of average functional activation changes with respect to various cortical regions.

In order to compare the accuracy of subject classification based on transfer task completion time or fNIRS-based metrics, several combinations of metrics are used for the classification models. These combinations include transfer task performance time only, and all possible combinations between PFC, LMM1, and SMA. Figure 4A shows

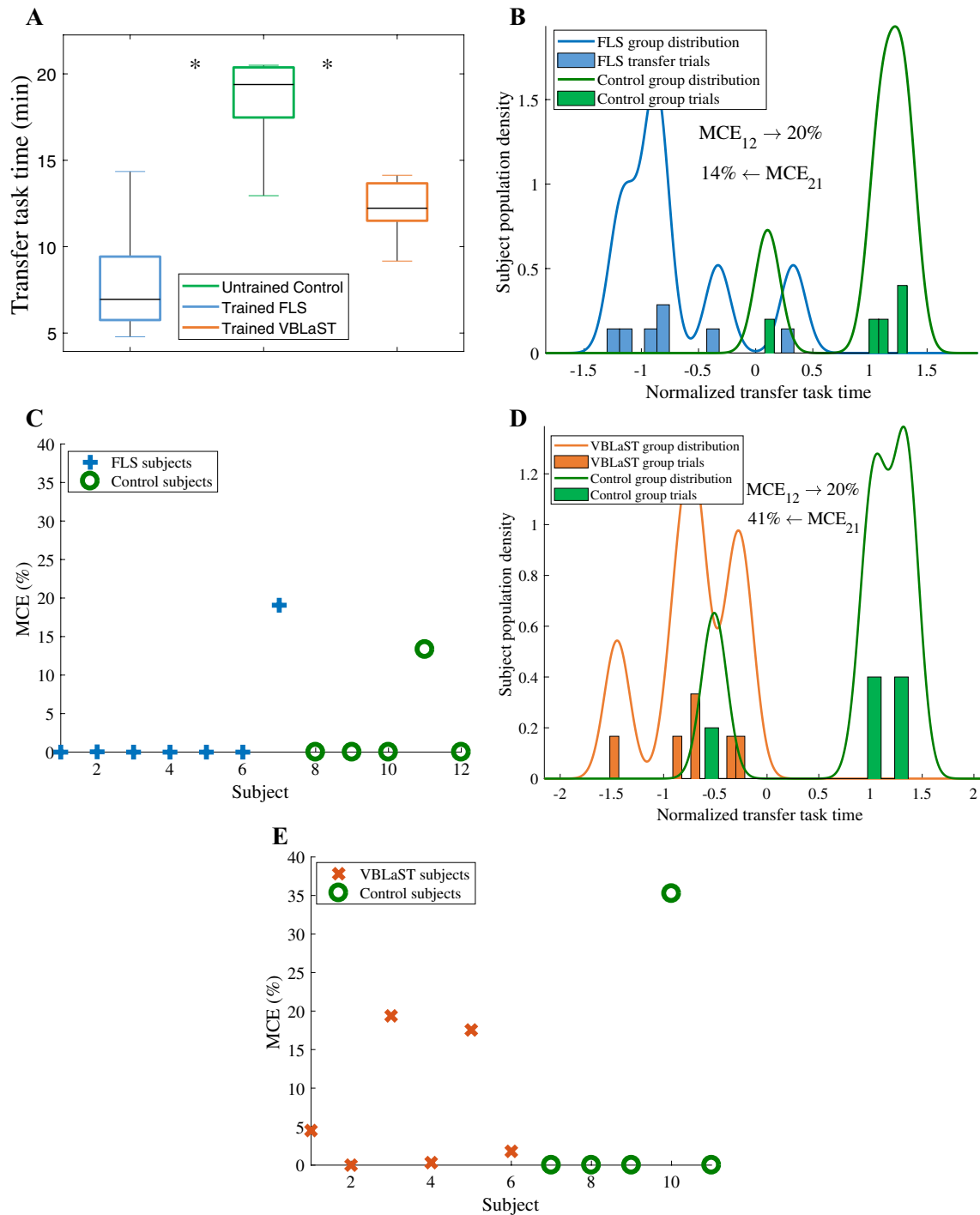


Fig. 2 Classification of surgical motor skill transfer based on task completion time. **A** Transfer task completion times for the trained FLS, untrained control, and trained VBLaST subjects ($*p < 0.05$). **B** LDA classification of trained FLS and control subjects during the

transfer task based on completion times and **C** corresponding cross-validation results. **D** LDA classification of trained VBLaST and control subjects during the transfer task based on completion times and **E** corresponding cross-validation results

the relative MCEs for various combinations of performance and fNIRS metrics to classify FLS trained subjects from untrained control subjects. The fNIRS metrics combination of PFC + LMM1 + SMA used for the FLS

classification model yields very low misclassification errors ($MCE_1 = 2.2\%$, $MCE_2 = 2.7\%$).

Similarly, the fNIRS metrics combination of PFC + LMM1 + SMA used for the VBLaST

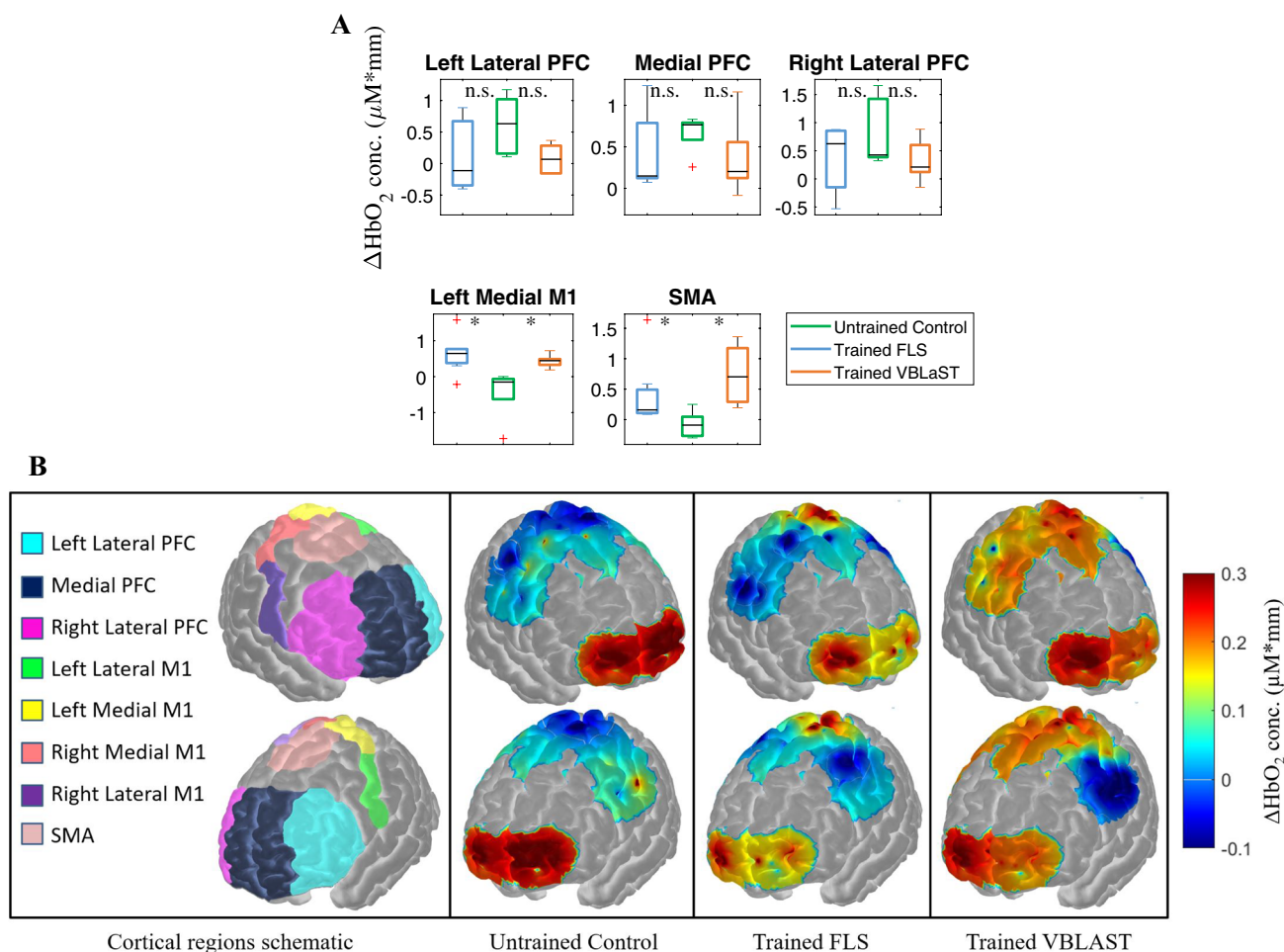


Fig. 3 Changes in cortical activation during the transfer task with respect to cortical regions. **A** Average changes in hemoglobin (ΔHbO_2) concentration as a measure of functional activation with respect to different cortical regions for FLS trained subjects (magenta), untrained control subjects (cyan), and VBLaST trained

subjects (black) while all subjects perform the ex-vivo transfer task. **B** Spatial maps of functional activation with respect to varying degrees of training levels during the ex-vivo transfer task. Spatial maps cover specific regions including the PFC, M1, and SMA for all subjects. (Color figure online)

classification model yields very low misclassification errors ($\text{MCE}_1 = 8.9\%$, $\text{MCE}_2 = 9.1\%$), as shown in Fig. 4B. Figure 4C shows the cross-validation results of various classification models to classify trained FLS or VBLaST subjects with untrained control subjects. FLS trained versus control subjects classification based on transfer performance scores and PFC + LMM1 + SMA combinations yield results where 83% of the samples have MCE less than 0.05. In a similar fashion, VBLaST trained vs control subjects classification models show that the transfer task performance score and PFC + LMM1 + SMA metric combinations yield in 72% of the samples with MCE less than 0.05. These cross-validation results independently assess the accuracy of both classification models using transfer task time and PFC + LMM1 + SMA metrics, ultimately showing that the resulting MCEs from both classification molds can be objectively compared.

Discussion

Accurate and objective assessment of surgical skills transfer from simulation environments to clinical settings is vital in determining the effectiveness of surgical training. Current standards utilizing rating checklists or task completion time metrics are limited in reliability, when objectively determining motor skill transfer to clinical environments [16–18, 52, 53]. For the first time, we present evidence that a neuroimaging-based approach provides objective assessment of surgical skill transfer from simulation to clinically relevant environments. The results are independent of whether the simulated task was in a physical or a virtual simulator and have been independently assessed to be robust in classifying trained and untrained subjects.

Note that the defacto metric used in numerous validation studies to show surgical skill transfer is performance time

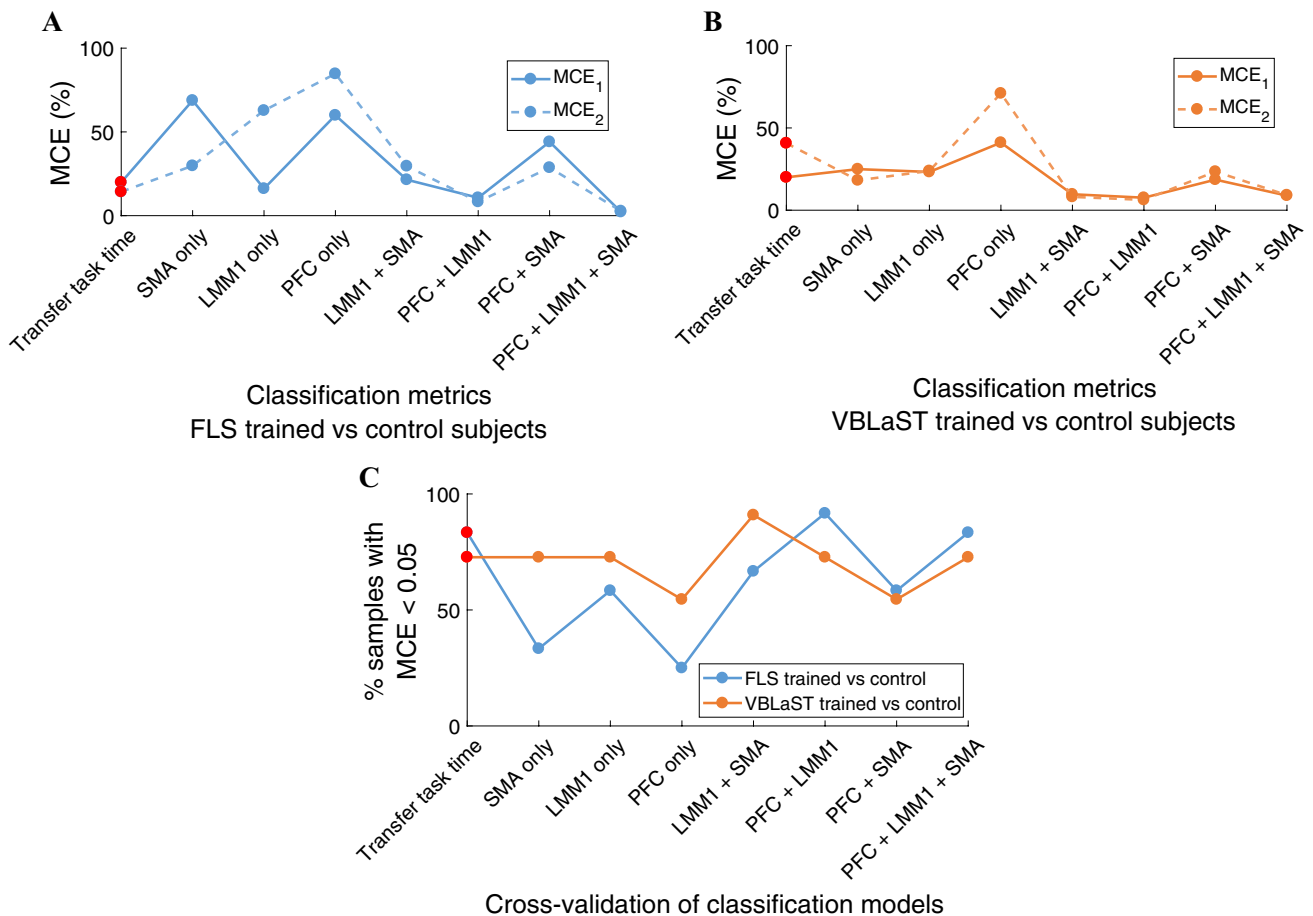


Fig. 4 **A** Classification results to classify FLS trained, VBLaST trained, and untrained control subjects. The cumulative set of MCE₁ and MCE₂ for all combinations of fNIRS metrics and the transfer task completion time to classify FLS trained and control subjects. MCE₁ indicates the probability that a trained subject is misclassified as an untrained control subject and MCE₂ indicates the probability that an untrained control subject is misclassified as a trained subject. **B** The

cumulative set of MCE₁ and MCE₂ for all combinations of fNIRS and transfer task metrics to classify VBLaST trained and control subjects. **C** Leave-one-out cross-validation results indicate the MCE for each sample treated as an independent sample for the LDA model using all combinations of fNIRS and transfer task metrics, where the percent of samples that have MCE below 0.05 for each possible metric combination are shown

[16]. While our results corroborate the notion that decreases in task performance time are features of expert surgical skills, utilizing this metric alone leads to inconsistencies in literature [16, 53–55]. This point is further supported by our classification models where task performance time metrics present 20–41% MCE indicating a lack of robustness. Since no single metric itself, such as task completion time, can demonstrate surgical skill proficiency between trained and untrained subjects [16, 54], our fNIRS metrics-based multivariate approach on classifying trained and untrained subjects brings robustness in surgical skill transfer assessment. Unfortunately, task quality measures are also subjective and not standardized for simulation paradigms, further prompting a need for alternative methods such as our neuroimaging-based approach [16–18, 54].

Using fNIRS as a means to measure functional brain activation in real time, we have shown that FLS and VBLaST

trained subjects show significant increases in activation in the left medial M1 and SMA, however, no significant differences in the PFC. These regions have been deliberately chosen due to their influence on motor task planning, execution, and fine motor control for complex motor tasks and their critical role in motor skill learning [38, 39, 56–60]. Specifically, the PFC is associated with motor strategy and the early stages of motor skill learning. The M1 and SMA are associated with execution and fine motor control and show increased activation during the later stages of motor skill learning as an indication of procurement of fine motor skills.

Our results are consistent with literature findings that indicate that subjects with fine motor skills in complex motor tasks exhibit higher M1 and SMA activation, particularly for bimanual motor tasks [41, 42, 61]. Furthermore, since all the subjects are right handed, majority of the fine motor manipulations employed during the pattern cutting

task is via the right hand. Since right-handed motor tasks evoke contralateral activation in the left hemisphere of the cortex, we expect increased activity in the left medial M1 [38, 39, 56–60]. Although we do not report any significant cortical activation differences between the untrained and trained subjects in the PFC during the transfer task, this is an expected result since all the subjects are expected to recruit the PFC to develop a motor strategy for this unfamiliar transfer task. While these results show promise in assessing surgical motor skill transfer via brain imaging techniques, future studies are required for a formative conclusion. These studies would prospectively include increased sample sizes and subject recruitment, high density probes for higher spatial resolution for imaging, and inclusion of other FLS tasks for transfer task assessment.

Using well-established neurophysiological principles, our work integrates most recent advances in neuroimaging and assessment of surgical competence during transfer of skills from a simulation environment. Since fNIRS signals are heavily contaminated by superficial tissue, short separation channel regression can be used to further isolate cortical brain activation signals from superficial tissue [49, 62]. Such approaches provide more robust estimations of the underlying hemodynamic responses associated with surgical tasks, which were not reported in previous fNIRS surgical studies.

Conclusion

Here, we propose fNIRS as a non-invasive real-time imaging method to successfully differentiate and classify surgical motor skills that transfer from simulation to ex-vivo environments. First, we show that conventional surgical skill transfer metrics, such as task completion time, have significantly high MCE, when used to classify trained and untrained subjects in assessing surgical motor skill transfer. We also show that fNIRS-based metrics have significantly lower MCE than task completion time for surgical skill transfer assessment. fNIRS-based approaches to objectively quantify motor skill transfer may be a paradigm change for the surgical community in determining the effectiveness of surgical trainers in training technical skills that ultimately transfer to the operating room.

Acknowledgements The authors would like to thank the medical student subjects and their dedication for this study. The authors would also like to thank the anatomical gift program and the gross anatomy lab at University at Buffalo for their support regarding the ex-vivo cadaveric samples. We would also like to thank Arthur “Buzz” DiMartino and his team at TechEn in graciously providing us support with the CW6 spectrometer.

Funding This work is supported by NIBIB 1R01EB014305, NHBLI 1R01HL119248, and NCI 1R01CA197491 grants that were awarded to Suvranu De.

Compliance with ethical standards

Disclosures Drs. Arun Nemani, Uwe Kruger, Meryem Yucel, Clairice Cooper, Steven Schwaizberg, Xavier Intes, and Suvranu De have no conflicts of interest or financial ties to disclose.

References

1. Dawe SR, Windsor JA, Broeders JAJL, Cregan PC, Hewett PJ, Maddern GJ (2014) A systematic review of surgical skills transfer after simulation-based training. *Ann Surg* 259:236–248. <https://doi.org/10.1097/SLA.0000000000000245>
2. Fraser SA, Klassen DR, Feldman LS, Ghitulescu GA, Stanbridge D, Fried GM (2003) Evaluating laparoscopic skills: Setting the pass/fail score for the MISTELS system. *Surg Endosc* 17:964–967. <https://doi.org/10.1007/s00464-002-8828-4>
3. Fraser SA, Feldman LS, Stanbridge D, Fried GM (2005) Characterizing the learning curve for a basic laparoscopic drill. *Surg Endosc* 19:1572–1578. <https://doi.org/10.1007/s00464-005-0150-5>
4. Fried GM, Feldman LS, Vassiliou MC, Fraser SA, Stanbridge D, Ghitulescu G, Andrew CG (2004) Proving the value of simulation in laparoscopic surgery. *Ann Surg* 240:518–525. <https://doi.org/10.1097/01.SLA.0000136941.46529.56>
5. McCluney AL, Vassiliou MC, Kaneva PA, Cao J, Stanbridge DD, Feldman LS, Fried GM (2007) FLS simulator performance predicts intraoperative laparoscopic skill. *Surg Endosc* 21:1991–1995. <https://doi.org/10.1007/s00464-007-9451-1>
6. Scott DJ, Ritter EM, Tesfay ST, Pimentel EA, Nagji A, Fried GM (2008) Certification pass rate of 100% for fundamentals of laparoscopic surgery skills after proficiency-based training. *Surg Endosc* 22:1887–1893. <https://doi.org/10.1007/s00464-008-9745-y>
7. Satava RM (2010) Emerging trends that herald the future of surgical simulation. *Surg Clin North Am* 90:623–633. <https://doi.org/10.1016/j.suc.2010.02.002>
8. Fried GM (2008) FLS assessment of competency using simulated laparoscopic tasks. *J Gastrointest Surg* 12:210–212. <https://doi.org/10.1007/s11605-007-0355-0>
9. Poulouse BK, Vassiliou MC, Dunkin BJ, Mellinger JD, Fanelli RD, Martinez JM, Hazey JW, Sillin LF, Delaney CP, Velanovich V, Fried GM, Korndorffer JR, Marks JM (2014) Fundamentals of endoscopic surgery cognitive examination: development and validity evidence. *Surg Endosc* 28:631–638. <https://doi.org/10.1007/s00464-013-3220-0>
10. Vassiliou MC, Dunkin BJ, Fried GM, Mellinger JD, Trus T, Kaneva P, Lyons C, Korndorffer JR, Ujiki M, Velanovich V, Kochman ML, Tsuda S, Martinez J, Scott DJ, Korus G, Park A, Marks JM (2014) Fundamentals of endoscopic surgery: creation and validation of the hands-on test. *Surg Endosc* 28:704–711. <https://doi.org/10.1007/s00464-013-3298-4>
11. McDougall EM (2007) Validation of surgical simulators. *J Endourol* 21:244–247. <https://doi.org/10.1089/end.2007.9985>
12. Scott DJ, Pugh CM, Ritter EM, Jacobs LM, Pellegrini CA, Sachdeva AK (2011) New directions in simulation-based surgical education and training: validation and transfer of surgical skills, use of nonsurgeons as faculty, use of simulation to screen and select surgery residents, and long-term follow-up of learners. *Surgery* 149:735–744. <https://doi.org/10.1016/j.surg.2010.11.010>
13. Aggarwal R, Grantcharov T, Moorthy K, Milland T, Papasavas P, Dosis A, Bello F, Darzi A (2007) An evaluation of the feasibility, validity, and reliability of laparoscopic skills assessment in the operating room. *Ann Surg* 245:992–999. <https://doi.org/10.1097/01.sla.0000262780.17950.e5>

14. Vassiliou MC, Feldman LS, Andrew CG, Bergman S, Leffondré K, Stanbridge D, Fried GM (2005) A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am J Surg* 190:107–113. <https://doi.org/10.1016/j.amjsurg.2005.04.004>
15. Doyle JD, Webber EM, Sidhu RS (2007) A universal global rating scale for the evaluation of technical skills in the operating room. *Am J Surg* 193:551–555. <https://doi.org/10.1016/j.amjsurg.2007.02.003>
16. Dawe SR, Pena GN, Windsor JA, Broeders JAJL, Cregan PC, Hewett PJ, Maddern GJ (2014) Systematic review of skills transfer after surgical simulation-based training. *Br J Surg* 101:1063–1076. <https://doi.org/10.1002/bjs.9482>
17. Moorthy K, Munz Y (2003) Objective assessment of technical skills in surgery. *Br Med J* 327:1032–1037. <https://doi.org/10.1136/bmj.327.7422.1032>
18. Hogle NJ, Chang L, Strong VEM, Welcome AOU, Sinaan M, Bailey R, Fowler DL (2009) Validation of laparoscopic surgical skills training outside the operating room: a long road. *Surg Endosc* 23:1476–1482. <https://doi.org/10.1007/s00464-009-0379-5>
19. Bosecker C, Dipietro L, Volpe B, Krebs HI (2010) Kinematic robot-based evaluation scales and clinical counterparts to measure upper limb motor performance in patients with chronic stroke. *Neurorehabil Neural Repair* 24:62–69. <https://doi.org/10.1177/1545968309343214>
20. Buckley CE, Kavanagh DO, Traynor O, Neary PC (2014) Is the skillset obtained in surgical simulation transferable to the operating theatre? *Am J Surg* 207:146–157. <https://doi.org/10.1016/j.amjsurg.2013.06.017>
21. Dai X, Zhang T, Yang H, Tang J, Carney PR, Jiang H (2017) Fast non-invasive functional diffuse optical tomography for brain imaging. *J Biophotonics*. <https://doi.org/10.1002/jbio.201600267>
22. Leff DR, Orihuela-Espina F, Elwell CE, Athanasiou T, Delpy DT, Darzi AW, Yang G-Z (2011) Assessment of the cerebral cortex during motor task behaviours in adults: a systematic review of functional near infrared spectroscopy (fNIRS) studies. *Neuroimage* 54:2922–2936. <https://doi.org/10.1016/j.neuroimage.2010.10.058>
23. Leff DR, Elwell CE, Orihuela-Espina F, Atallah L, Delpy DT, Darzi AW, Yang GZ (2008) Changes in prefrontal cortical behaviour depend upon familiarity on a bimanual co-ordination task: an fNIRS study. *Neuroimage* 39:805–813. <https://doi.org/10.1016/j.neuroimage.2007.09.032>
24. James DRC, Orihuela-Espina F, Leff DR, Sodergren MH, Athanasiou T, Darzi AW, Yang G-Z (2011) The ergonomics of natural orifice transluminal endoscopic surgery (NOTES) navigation in terms of performance, stress, and cognitive behavior. *Surgery* 149:525–533. <https://doi.org/10.1016/j.surg.2010.11.019>
25. Ohuchida K, Kenmotsu H, Yamamoto A, Sawada K, Hayami T, Morooka K, Takasugi S, Konishi K, Ieiri S, Tanoue K, Iwamoto Y, Tanaka M, Hashizume M (2009) The frontal cortex is activated during learning of endoscopic procedures. *Surg Endosc* 23:2296–2301. <https://doi.org/10.1007/s00464-008-0316-z>
26. Shewokis PA, Shariff FU, Liu Y, Ayaz H, Castellanos A, Lind DS (2017) Acquisition, retention and transfer of simulated laparoscopic tasks using fNIR and a contextual interference paradigm. *Am J Surg* 213:336–345. <https://doi.org/10.1016/j.amjsurg.2016.11.043>
27. Andreu-Perez J, Leff DR, Shetty K, Darzi A, Yang G-Z (2016) Disparity in frontal lobe connectivity on a complex bimanual motor task aids in classification of operator skill level. *Brain Connect* 6:375–388. <https://doi.org/10.1089/brain.2015.0350>
28. Modi HN, Singh H, Yang G-Z, Darzi A, Leff DR, Mansur A, Yang AE, Darzi G-Z, Leff A DR (2017) A decade of imaging surgeons' brain function (part I): terminology, techniques, and clinical translation. *Surgery* 28:2189–2198. <https://doi.org/10.1016/j.surg.2017.05.021>
29. Nemani A, Kruger U, Intes X, De S (2017) Increased sensitivity in discriminating surgical motor skills using prefrontal cortex activation over established metrics. In: Optics in the life sciences congress. OSA, San Diego, p 11
30. Soper NJ, Fried GM (2008) The fundamentals of laparoscopic surgery: its time has come. *Bull Am Coll Surg* 93:30–32
31. Peters JH, Fried GM, Swanstrom LL, Soper NJ, Sillin LF, Schirmer B, Hoffman K, the SAGES FLS Committee the SF (2004) Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. *Surgery* 135:21–27. [https://doi.org/10.1016/S0039-6060\(03\)00156-9](https://doi.org/10.1016/S0039-6060(03)00156-9)
32. Zhang L, Sankaranarayanan G, Arikatla VS, Ahn W, Grosdemouge C, Rideout JM, Epstein SK, De S, Schwaizberg SD, Jones DB, Cao CGL (2013) Characterizing the learning curve of the VBLaST-PT(©) (Virtual basic laparoscopic skill trainer). *Surg Endosc* 27:3603–3615. <https://doi.org/10.1007/s00464-013-2932-5>
33. Maciel A, Liu Y, Ahn W, Singh TP, Dunnican W, De S (2008) Development of the VBLaST: a virtual basic laparoscopic skill trainer. *Int J Med Robot Comput Assist Surg* 4:131–138. <https://doi.org/10.1002/rcs.185>
34. Arikatla VS, Sankaranarayanan G, Ahn W, Chellali A, De S, Caroline GL, Hwabejire J, DeMoya M, Schwaizberg S, Jones DB (2013) Face and construct validation of a virtual peg transfer simulator. *Surg Endosc* 27:1721–1729. <https://doi.org/10.1007/s00464-012-2664-y>
35. Sankaranarayanan G, Lin H, Arikatla VS, Mulcare M, Zhang L, Derevianko A, Lim R, Fobert D, Cao C, Schwaizberg SD, Jones DB, De S (2010) Preliminary face and construct validation study of a virtual basic laparoscopic skill trainer. *J Laparoendosc Adv Surg Tech* 20:153–157. <https://doi.org/10.1089/lap.2009.0030>
36. Chellali A, Ahn W, Sankaranarayanan G, Flinn JT, Schwaizberg SD, Jones DB, De S, Cao CGL (2015) Preliminary evaluation of the pattern cutting and the ligating loop virtual laparoscopic trainers. *Surg Endosc* 29:815–821. <https://doi.org/10.1007/s00464-014-3764-7>
37. Nemani A, Ahn W, Cooper C, Schwaizberg S, De S Convergent validation and transfer of learning studies of a virtual reality-based pattern cutting simulator. *Surg Endosc*. <https://doi.org/10.1007/s00464-017-5802-8>
38. Hikosaka O, Nakamura K, Sakai K, Nakahara H (2002) Central mechanisms of motor skill learning. *Curr Opin Neurobiol* 12:217–222. [https://doi.org/10.1016/S0959-4388\(02\)00307-0](https://doi.org/10.1016/S0959-4388(02)00307-0)
39. Wolpert DM, Diedrichsen J, Flanagan JR (2011) Principles of sensorimotor learning. *Nat Rev Neurosci* 12:739–751. <https://doi.org/10.1038/nrn3112>
40. Swinnen SP (2002) Intermanual coordination: from behavioural principles to neural-network interactions. *Nat Rev Neurosci* 3:348–359. <https://doi.org/10.1038/nrn807>
41. Swinnen SP, Gooijers J (2015) Bimanual Coordination. In: Brain mapp, 1st edn. Elsevier, San Diego, pp 475–482
42. Swinnen SP, Wenderoth N (2004) Two hands, one brain: cognitive neuroscience of bimanual skill. *Trends Cogn Sci* 8:18–25
43. Nemani A, Intes X, De S (2014) Monte Carlo based simulation of sensitivity curvature for evaluating optimal probe geometry. In: Biomedical optics 2014. OSA, Washington, D.C., p BM3A.36
44. Cope M, Delpy DT (1988) System for long-term measurement of cerebral blood and tissue oxygenation on newborn infants by near infra-red transillumination. *Med Biol Eng Comput* 26:289–294. <https://doi.org/10.1007/BF02447083>
45. Delpy DT, Cope M, Zee P, van der Arridge S, Wray S, Wyatt J (1988) Estimation of optical pathlength through tissue from direct time of flight measurement. *Phys Med Biol* 33:1433–1442. <https://doi.org/10.1088/0031-9155/33/12/008>

46. Huppert TJ, Diamond SG, Franceschini MA, Boas DA (2009) Homer: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *Appl Opt* 48:280–298. <https://doi.org/10.1364/AO.48.00D280>
47. Franceschini MA, Joseph DK, Huppert TJ, Diamond SG, Boas DA (2006) Diffuse optical imaging of the whole head. *J Biomed Opt* 11:1–22. <https://doi.org/10.1117/1.2363365>
48. Zhang Y, Brooks DH, Franceschini MA, Boas DA (2005) Eigenvector-based spatial filtering for reduction of physiological interference in diffuse optical imaging. *J Biomed Opt* 10:1–11. <https://doi.org/10.1117/1.1852552>
49. Gagnon L, Yücel MA, Boas DA, Cooper RJ (2014) Further improvement in reducing superficial contamination in NIRS using double short separation measurements. *Neuroimage* 85:127–135. <https://doi.org/10.1016/j.neuroimage.2013.01.073>
50. Tabachnick BG, Fidell LS (2007) Using multivariate statistics, 5th edn. Allyn & Bacon, Needham Heights
51. Silverman BW (1986) Density estimation for statistics and data analysis. CRC press, Boca Raton
52. Ahlberg G, Enochsson L, Gallagher AG, Hedman L, Hogman C, McClusky DA, Ramel S, Smith CD, Arvidsson D (2007) Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. *Am J Surg* 193:797–804. <https://doi.org/10.1016/j.amjsurg.2006.06.050>
53. Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, Satava RM (2002) Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg* 236:458–463. <https://doi.org/10.1097/01.SLA.0000028969.51489.B4>
54. Ahlberg G, Hultcrantz R, Jaramillo E, Lindblom A, Arvidsson D (2005) Virtual reality colonoscopy simulation: a compulsory practice for the future colonoscopist? *Endoscopy* 37:1198–1204. <https://doi.org/10.1055/s-2005-921049>
55. Grantcharov TP, Kristiansen VB, Bendix J, Bardram L, Rosenberg J, Funch-Jensen P (2004) Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *Br J Surg* 91:146–150. <https://doi.org/10.1002/bjs.4407>
56. Sakai K, Hikosaka O, Miyauchi S, Takino R, Sasaki Y, Pütz B (1998) Transition of brain activation from frontal to parietal areas in visuomotor sequence learning. *J Neurosci* 18:1827–1840
57. Honda M (1998) Dynamic cortical involvement in implicit and explicit motor sequence learning: a PET study. *Brain* 121:2159–2173. <https://doi.org/10.1093/brain/121.11.2159>
58. Nakamura K, Sakai K, Hikosaka O (1998) Neuronal activity in medial frontal cortex during learning of sequential procedures. *J Neurophysiol* 80:2671–2687
59. Klintsova AY, Greenough WT (1999) Synaptic plasticity in cortical systems. *Curr Opin Neurobiol* 9:203–208
60. Wolpert DM, Ghahramani Z, Flanagan JR (2001) Perspectives and problems in motor learning. *Trends Cogn Sci* 5:487–494. [https://doi.org/10.1016/S1364-6613\(00\)01773-3](https://doi.org/10.1016/S1364-6613(00)01773-3)
61. Serrien DJ, Ivry RB, Swinnen SP (2006) Dynamics of hemispheric specialization and integration in the context of motor control. *Nat Rev Neurosci* 7:160–166. <https://doi.org/10.1038/nrn1849>
62. Gagnon L, Cooper RJ, Yücel MA, Perdue KL, Greve DN, Boas DA (2012) Short separation channel location impacts the performance of short channel regression in NIRS. *Neuroimage* 59:2518–2528. <https://doi.org/10.1016/j.neuroimage.2011.08.095>